



The Patent Office
Concept House
Cardiff Road
Newport

South Wales
NP10 8QQ

WIPO PCT

GB99 11508

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

**PRIORITY
DOCUMENT**

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

Signed

AmBrewer

Dated 10 June 1999

THIS PAGE BLANK (USPTO)

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

12 MAY 1998

1. Your reference

M98/0209/GB

2. Patent application number

(The Patent Office will fill in this part)

3. Full name, address and postcode of the or of each applicant (*underline all surnames*)

University of Manchester Institute
of Science and Technology
P O Box 88
Manchester
M60 1QD

Patents ADP number (*if you know it*)

If the applicant is a corporate body, give the
country/state of its incorporation

Great Britain

4. Title of the invention

Visualising Images

5. Name of your agent (*if you have one*)

McNeight & Lawrence
Regent House
Heaton Lane
Stockport
Cheshire SK4 1BS

"Address for service" in the United Kingdom
to which all correspondence should be sent
(*including the postcode*)

Patents ADP number (*if you know it*)

0001115001

6. If you are declaring priority from one or more
earlier patent applications, give the country
and the date of filing of the or of each of these
earlier applications and (*if you know it*) the or
each application number

Country

Priority application number
(*if you know it*)

Date of filing
(*day / month / year*)

7. If this application is divided or otherwise
derived from an earlier UK application,
give the number and the filing date of
the earlier application

Number of earlier application

Date of filing
(*day / month / year*)

8. Is a statement of inventorship and of right
to grant of a patent required in support of
this request? (*Answer 'Yes' if:*

Yes

- a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an
applicant, or
c) any named applicant is a corporate body.
See note (d))

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form

Description 15

Claim(s)

Abstract

Drawing(s) 5

10. If you are also filing any of the following, state how many against each item.

Priority documents

Translations of priority documents

Statement of inventorship and right to grant of a patent (*Patents Form 7/77*)

Request for preliminary examination and search (*Patents Form 9/77*)

Request for substantive examination (*Patents Form 10/77*)

Any other documents
(please specify)

11. I/We request the grant of a patent on the basis of this application.

Signature

Date 11.05.98

McNeight & Lawrence

12. Name and daytime telephone number of person to contact in the United Kingdom

Dr Ian R Lambert

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

- a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.
- b) Write your answers in capital letters using black ink or you may type them.
- c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.
- d) If you have answered 'Yes' Patents Form 7/77 will need to be filed.
- e) Once you have filled in the form you must remember to sign and date it.
- f) For details of the fee and ways to pay please contact the Patent Office.

VISUALISING IMAGES

This invention relates to methods and devices enabling a person to visualise images.

The prospect of enabling subjects to visualise images through some external means, circumventing the human visual system, is one of clear significance. In particular, such a system might enable blind persons to "see", or at least assimilate some amount of visual information. Although systems based on echo location and on touch are known, there is at present no available system permitting images to be analysed in detail.

It is known that there are areas of similarity in the way the human brain interprets visual and auditory information. As with vision, auditory information is partitioned into discrete packages and conveyed to the relevant brain areas for separate processing : human speech sounds such as words and phrases are processed by Wernicke's area in the left hemisphere, where music is processed in the temporal lobe of the right hemisphere. In some animals, there are areas of the brain concerned with spatial perception in which both the visual and auditory topographical maps (involved in the location of objects in the environment by vision or by hearing) are superimposed. However, a deeper explanation of such similarities, i.e. the fundamental mechanisms underlying such similarities, is not, at present, available.

The present invention is based on the surprising discovery that musical forms can be used to convey precise visual information to a subject. Such "precise" information can comprise spatial information such as the precise shapes of objects or symbols, and should be delineated from imagery which can be evoked in subjects listening to a favourite piece of music, in which instance images "brought to mind" by the music are personal in nature, and can vary quite dramatically from subject to subject.

Almost certainly, this surprising discovery is related to the fundamental mechanisms governing the way in which the human brain segments, organises and processes information from various sources in multi dimensional space. However, discussion of such mechanisms is not the purpose of the present application.

According to a first aspect of the invention there is provided a method enabling a person to visualise images comprising the steps of :

encoding spatial information relating to a feature or features contained within an image into the form of one or more musical sequences; and

playing the musical sequence or sequences to the person.

“Spatial information” includes the shape, size, orientation and relative positions of features, as well as finer details such as surface decoration or, for example, the appearance of a face. As will be explained in more detail below, further visual information, such as colour and brightness, and temporal information, i.e. the movement of features, may also be visualised using the present invention. Features can be, for example, three dimensional objects, two dimensional objects such as drawings, or symbols such as letters, words and numbers.

Spatial information may be encoded by:

representing the image as a two dimensional (2D) image; and

forming one or more musical sequences, each comprising a series of notes or chords, in which i) each note or chord is selected dependant upon the distribution of the feature or features along a portion of the 2D image and ii) different notes or chords in a sequence correspond to different portions of the 2D image.

The 2D image, or a portion of the 2D image, may be divided into a matrix of pixels, and i) each note or chord may be selected dependent upon the distribution of the feature or features along a column (or row) of pixels and ii) different notes or chords in a sequence may correspond to the distribution of the feature or feature along different columns (or rows) of pixels. A different note may be associated with each pixel along a column and, if a feature is recognised as being present in a pixel, the note corresponding to that pixel comprises part of the musical sequence.

The method may enable a person to visualise moving features, and may comprise the step of playing a plurality of musical sequences corresponding to different positions and/or orientating of the moving feature.

The image may be encoded into the form of a plurality of musical sequences which are played to the person as a melody.

The image may be encoded as a plurality of musical sequences, each corresponding to different spatial resolutions. The image may be divided into two or more concentric zones, the zone at the centre of the image being encoded at the highest spatial resolution and the zone furthest from the centre of the image being encoded at the lowest spatial resolution. A feature or features may be visualised by obtaining a plurality of images in a saccadic movement. An example is the visualisation of a face, in which features such as eyes, nose and mouth are "scanned" at high resolution in a saccadic movement which mimics the operation of the human retina.

The spatial resolution corresponding to a musical sequence may be indicated by the duration of the notes and/or chords in the sequence.

The colour of the feature or features may be encoded by producing a musical sequence or sequences which comprise a plurality of different waveforms mixed

in variable ratios, the waveforms being selected so that none of the waveforms may be created by a linear combination of the other two waveforms. Three waveforms may be mixed in variable ratios. The three waveforms, may be produced by filtering a master waveform between different frequency ranges.

The brightness of the feature or features may be encoded by varying the intensity of the musical sequence or sequences, and/or by varying the pitch of one or more notes or chords.

According to a second aspect of the invention there is provided a device enabling a person to visualise images comprising:

imaging means for obtaining images of a feature or features;

encoding means for encoding spatial information relating to the feature or features according to the first aspect of the invention; and

playing means for playing the musical sequence or sequences to the person.

The imaging means may comprise at least one video camera.

The encoding means may comprise a microprocessor.

The playing means may comprise an ear piece.

The device may be portable, in which instance the imaging means may be hand-held or disposed on the person's head.

Methods and devices in accordance with the invention will now be described with reference to the accompanying drawings, in which:-

- Figure 1 shows a first 2D feature;
- Figure 2 shows a second 2D feature;
- Figure 3 shows a sequence of locomotory motions of a bird;
- Figure 4 shows the division of an image field into a number of zones of different resolution; and
- Figure 5 shows the coupling of different musical sequences in a melody.

The invention comprises a method enabling a person to visualise images comprising the steps of:

encoding spatial information relating to a feature or features contained within an image into the form of one or more musical sequences; and

playing the musical sequence or sequences to the person.

Figure 1 illustrates how the encoding may be performed on a computer generated 2D feature 10. The image 10 is, essentially, an arrangement of filled pixels which have been selected from a matrix of pixels. The musical sequence is produced by associating a different note with each pixel along a column of pixels and, if a feature is recognised as being present in a pixel, the note corresponding to that pixel comprises part of the musical sequence. If a feature occupies more than one pixel in a given column

then a number of notes are played simultaneously, producing a chord. The entire image is encoded by performing this procedure for each column of pixels, thereby producing a sequence of notes or chords.

In Figure 1, this encoding procedure is performed using a moveable cursor 12. The cursor 12 is divided into 32 segments corresponding to the notes in four octaves of the scale of C major (which comprises, in ascending order, the eight notes C D E F A B C). The cursor 12 defines a Y axis. Encoding proceeds by moving the cursor 12 along the X axis, from left to right as viewed in Figure 1. Each movement of the cursor 12 samples a new column of pixels. If the cursor 12 encounters one or more filled pixels (corresponding to a portion of the feature 10) then the note or notes corresponding to the segments of the cursor 12 which have encountered the filled pixels are played. Thus, if the cursor 12 is moving across the screen at a velocity of p columns of pixels per second, the time between the playing of successive notes or chords is $1/p$ seconds.

In other words, spatial information corresponding to the shape of the figure in Cartesian coordinates is encoded, or transposed, into a musical sequence in which the Y ordinate is represented by musical notes and the X ordinate by time.

Returning to the specific example shown in Figure 1, it can be seen that movement of the cursor 12 over the figure 10 will result in the playing of a musical sequence in which the ascending notes G, A, B, C of the second lowest octave and C, D, E, F of the second highest octave are played in succession.

As a further example, Figure 2 shows a computer generated square figure 20 which also comprises a number of filled pixels. When the cursor 12 is moved across the square 20, the first component of the musical sequence is a chord which comprises the ten notes A, B, C of the second lowest octave and C, D, E, F, G, A, B of the second highest octave. The number of notes involved results in a chord which gives the

impression of density or thickness. The next eight components of the musical sequence are chords in which only two notes, A and B, are played, these notes corresponding to the top and bottom sides of the square 20. The result is a sound which might be described as "thinner". The final component of the musical sequence is the chord comprising the ten notes.

The use of the scale of C major (which does not contain flats or sharps) is not limiting : other musical scales may be used. Indeed, since the four octaves utilised in the above examples represents, approximately, the range of human hearing, thus limiting the Y axis resolution of the encoded image, it may be advantageous to utilise the chromatic scale. In principle, the image might be encoded using a different coordinate system than Cartesian coordinates, such as polar coordinates. However, there are strong neurological reasons for expecting that the human brain would be drastically less suited to assimilate information encoded in this manner.

A computer program was written in the C++ language, running under the Windows (RTM) operating system to enable 2D shapes and objects to be encoded using the approach described above. A Musical Instrument Digital Interface (MIDI) allowed interfacing to a sound card in order to play the musical sequence. Confidential tests were performed, using the software, on a number of blind subjects, and on (blind folded) sighted subjects. Extremely favourable results were obtained in tests which employed a variety of geometric shapes and letters. For example, subjects were quickly able to read simple words, having been trained on individual letters. Furthermore, subjects were able to recognise figures consisting of one geometric shape contained within another shape, (such as a triangle within a square), having been trained on the individual component geometric shapes.

In a further development, it is possible to visualise moving features by playing a plurality of musical sequences corresponding to different positions and/or

orientations of the moving feature. In this way, dynamical information can be visualised in a way which bears similarities to the principles of cinematography, in which successive frames showing different stages of the movement are shown.

The computer program described above was adapted to produce a series of images which simulate the locomotory sequence of limb movements displayed by a variety of animal species, namely i) a galloping horse; ii) a running cheetah, iii) a walking man; iv) a flying bird; v) a swimming fish; vi) a bipedal lizard; vii) a quadrupedal lizard; viii) a wriggling worm; and ix) a crawling locust. Figure 3 shows a sequence of images containing the feature of a flying bird.

Successive musical sequences, corresponding to different "frames" in sequences of images such as that shown in Figure 3, were played to subjects in the confidential tests. Subjects were able to distinguish between the different locomotory motions and, in some instances, were able to correctly identify a locomotory motion with no previous training using the locomotory motion.

It is, of course, possible to convey the motion of a feature across a scene, instead of, or as well as indicating relative motion of components such as limbs and wings.

We now turn to the problem of encoding more complex images. The approach adopted is to substantially mimic, in a number of aspects, the operation of the human eye. Figure 4 shows how an image might be divided into a number of concentric domains or zones 40, 42, 44, 46, for the purpose of encoding an image. The spatial resolution, i.e. the size of the pixels used to produce a musical sequence, is different in each zone 40, 42, 44, 46. More specifically, the image defined by the largest zone 40 is encoded at the lowest resolution, zone 42 corresponds to a medium resolution, zone 44 corresponds to a medium/high resolution, whilst the image defined by the smallest zone

46 is encoded at the highest resolution. The use of four zones in this way is similar to the structural divisions of the fundus of the human eye, in which visual activity is highest at the foveola, in the centre of the retina, and diminishes going radially from the centre of the retina, through the foveola, fovea, parafovea and perifovea. The use of four concentric zones is not limiting : different numbers of zones might be employed. In principle, the zones need not be concentric, although, for reasons outlined below, this configuration is strongly preferred.

In the present example, the Y ordinate in each zone is divided into 32 segments, corresponding to four octaves in the scale of C major, and the X ordinate in each zone is divided into 8 segments, corresponding to eight separate notes or chords in each musical sequence. Thus, the spatial resolution in zone 46 is higher than in zone 40 by virtue of the smaller dimensions of zone 46. An important aspect of this approach is that the zone represented by a musical sequence can be identified by virtue of the timing of the constituent notes or chords. More specifically, the higher the resolution, the faster the sequence is played. This exploits the fact the brain organises and segments visual and audio information in similar ways, i.e. notes sounded in close temporal proximity tend to be grouped together, in the same way that markings or features that are close together (and therefore require high spatial resolution to study them) tend to be grouped together.

In the present example, the musical notes or chords comprising each musical sequence are played at the following musical time intervals:

largest zone 40, lowest resolution	:	8 minims
second largest zone 42, medium resolution	:	8 crotchets
second smallest zone 44, medium/high resolution	:	8 quavers
smallest zone 46, high resolution	:	8 semiquavers

Thus the X axis is defined by musical time values. Although the Y axis is defined by a four octave musical scale, it can be useful to define the Y axis in terms of effective time values. The zone 40 can be considered as comprising a 8 x 8 minims region, zone 42 as comprising a 8 x 8 crotchet region, etc. In this instance we can define an effective pixel size. Thus, for example, the Y dimension of a pixel in zone 40 is 1/32 of 8 minims, or 1 quaver, in effective time units. Therefore, the spatial resolution of the zones can be expressed as the following pixel dimensions, where the first dimension relates to the X ordinate and the second dimension relates to the Y ordinate

low resolution	:	1 minim x 1 quaver
medium resolution	:	1 crotchet x 1 semiquaver
medium/high resolution	:	1 quaver x 1 demisemiquaver
high resolution	:	1 semiquaver x 1 hemidemisemiquaver

Figure 4 also shows the partial tiling of zone 46 by a plurality of tiles 48. The tiling extends from the centre of the field (marked with a cross) along a diagonal to the upper corner of the top left quadrant. The tiling pattern would then be continued along each of the corresponding diagonals in the other three quadrants and along X and Y axis passing through the centre of the field, thus completely tiling zone 46. The tiles can be used as neurons in an image recognition stage which may be required to convert an incoming image into a number of matrices of pixels. More specifically, tiling may be used to determine if a figure lies wholly within the area bounded by a given zone, or if a portion of the figure lies outside of the zone. It may be advantageous to restrict the musical sequence to encode only those features wholly contained within the zone. Such a process bears substantial similarities with the end-stopped cells found in the mammalian visual cortex. A further level of sophistication can be provided by using detectors, or image recognition techniques, which are adapted to respond to certain complex features of an image, such as orientation, corners or even written words. In the

present invention, the detection of such a feature would cause a pixel to release a spatiotemporally coded musical sequence at an appropriate time.

It will be apparent that it is possible to observe a scene at low resolution, and then to zoom in onto individual features at high resolution. Furthermore, the configuration of Figure 4 might be used to view a scene or object in saccadic fashion. Saccadic eye movements are rapid, ballistic movements of the eyes used in scanning a scene or object. An example is facial recognition in which the eye motion rapidly and successively puts features such as eyes, nose and mouth in the central, high resolution foveola and fovea zones of the retina. Due to time taken in playing a musical sequence, such saccadic movements using the present invention will be less rapid than in the human eye. However, it is quite feasible that a low resolution image, encoded using the present invention, might be used to indicate features of interest which are successively brought into the central portion of the image for visualisation at high resolution.

From the foregoing, it will be apparent that image or images can be segregated into a plurality of musical sequences, corresponding, for example, views at different resolutions, saccadic compilations of several related images, and "special" sequences relating to certain "programmed" features. These separate auditory representations may be bound into a single percept by the use of melody. Consider an example in which the outline of a head is presented on one sequence, followed by the finer details of the face (eyes, nose etc).

The image of the head would then be encoded in a (repeated) two-bar musical melody in which the outline of the head featured as the contents of one window, and the eyes etc contained in the other.

To generate the "melody" (which encompasses the contents of both windows) and to preserve a sense of the "spatial" relationships between them we introduce two additional modifications:

(i) Figure 5a illustrates how to alternate the presentations by modulating the sound intensity of each window presentation 50, 52 above and below the auditory threshold and arranging for the intensity modulation of the two presentations 50, 52 to be 180-degrees out of phase.

(ii) Figure 5b shows the continued modulation of the sound intensity of presentation 50, symmetrically about the auditory threshold. For the other presentation 52 we shift the baseline of its modulation a fraction up, so that it lies some distance above the auditory threshold. Thus we have introduced:

- (a) a tempo which musically brackets the contents of the two windows within the simple melody
- (b) a short interval at the start and end of each cycle when a portion of each window may be heard simultaneously - one rising in intensity, the other decreasing.

Thus by simply shifting the modulation baseline of one or the other above or below the auditory threshold one may selectively present and listen to:

- (i) the contents of one of the "voices".
- (ii) generate a full alternation of the contents of each window.
- (iii) a tempo that brackets the contents of the two windows: allowing the observer to perceive them as a single entity (much as the theme tune of a song).

- (iv) portions of both windows simultaneously.

This approach can be extended using principles well established in musicology in order to generate and modulate complex melodies with many subsidiary "voices". Each "voice" contains visual information, and furthermore, visual information can be contained in the relationships between the "voices". The individual musical sequences might utilise different waveforms, i.e. different instruments or different voices might be allocated to different musical sequences, giving rise to considerations of harmony.

The colour of features can be represented using the present invention. In classical colour theory, the perception of colour is generated through the differential absorption of different wavebands of light by the visual pigments contained in 3 types of photoreceptor which serve as primaries. Any 3 coloured lights can serve as primaries provided only that when mixed together in suitable proportions they produced the sensation of "white" and perhaps more importantly: on condition that it should not be possible to match one of those by linear combination of the other two.

One way of achieving such colour mixing with the present invention is to select a master waveform corresponding to a musical instrument which spans a reasonable range of octaves and whose notes contain a rich range of harmonic overtones. A triad of primary waveforms, for every notes in the span of octaves employed in the musical sequences, is generated as follows:

- (i) To generate the "long wavelength" version of that note we use a sound bandpass filter to remove some of the medium and higher frequency components from the sound normally generated by that note.

(ii) For the "medium wavelength" version of the same note we follow the same procedure except that we filter out some of the high and low frequency components.

(iii) For the "short wavelength" version we filter out some of the low and medium frequency components.

(iv) Next we adjust the relative intensities of the triad generated so that when they are sounded together they retrieve the sound generated by the original note (which serves as the achromatic note in the set).

(v) This triad should satisfy the principal condition required of a set of colour primaries: namely that it should not be possible for the sound of any one of the triad to be matched by a suitable (intensity) mixture of the other two.

The procedure described above is repeated for each note in the entire set of octaves utilised by the present invention. However to obtain a suitable set of triads for each note, the characteristics of the bandpass filters need to be altered in each case to take account of the change in pitch of the zero harmonic (fundamental) as we progress up or down the scale of notes.

Alternatively, it might be possible to employ three different instruments as the three colour primaries, i.e. the basis set from which colour is derived.

Another consideration still is the brightness of features in an image. Both normal (and synesthetic subjects) describe that the sound of notes gets brighter as we move towards the treble end of the scale and darker as we proceed towards the bass end. Synesthetic subjects also describe colours as becoming lighter as notes are sounded at the treble end and darker towards the bass end. Thus an ascending scale of notes will

appear to get brighter as we move towards the treble end. This suggests it may be necessary to introduce some other musical artifice. One possibility is an additional musical instrument, suitably calibrated to provide pitch to signal brightness, as a background accompaniment. Another possibility is to modulate the intensity of the relevant musical sequence or sequences in order to indicate brightness.

The present invention provides devices enabling a person to visualise images. Such devices comprise: imaging means for obtaining images of a feature or features; encoding means for encoding spatial information relating to the feature or features in the manner described above; and playing means for playing the musical sequence or sequences to the person.

The imaging means can comprise a video camera, although other means, such as CCD detectors, might be employed. The encoding means performs the functions of analysing the image produced by the imaging means in a suitable manner, and encoding the analysed image into suitable musical sequences. The analysis step might comprise the division of the image, or portions of the image, into the desired number of pixels. It is of course highly desirable that the device is portable, and thus a small, dedicated microprocessor might be used as encoding means. A small video camera can be used as part of a portable device: the video camera can be incorporated into a hand-held "wand", or positioned on the person's head, possibly with a head band, in the manner of a headlamp. In both instances, saccadic movements can be accomplished by the period via hand or head motion. The playing means can comprise an ear-piece worn by the person.

This Page Blank (uspto)

C
B
A
G
F
E
D
C
C
C
B
A
G
F
E
D
C
C
B
A
G
F
E
D
C
C
B
A
G
F
E
D
C

12

10

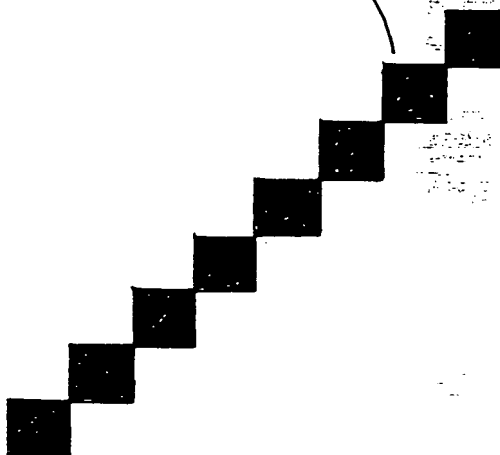


Fig 1

This Page Blank (uspto)

C
B
A
G
F
E
O
C
C
B
A
G
F
E
D
C
C
C
B
A
G
F
E
D
C
C
B
A
G
F
E
D
C

12

20

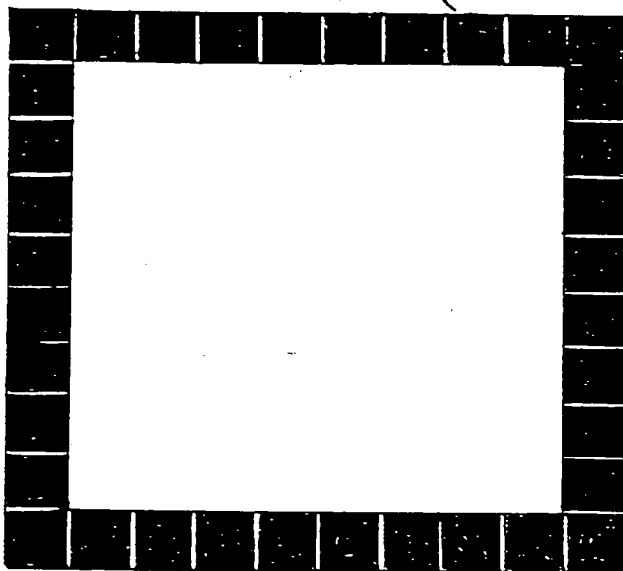
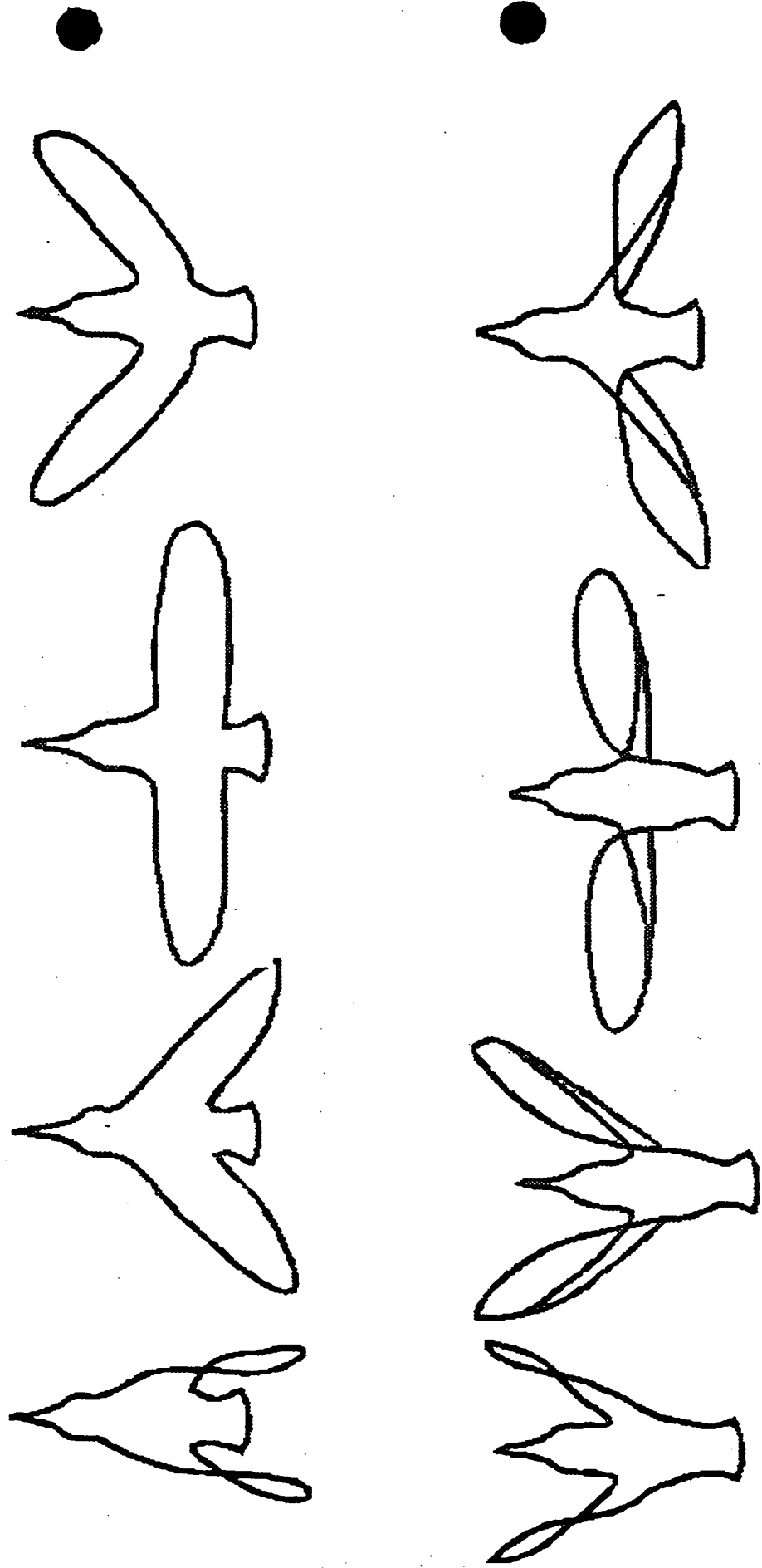


Fig. 2

This Page Blank (uspto)

Fig. 3



This Page Blank (uspto)

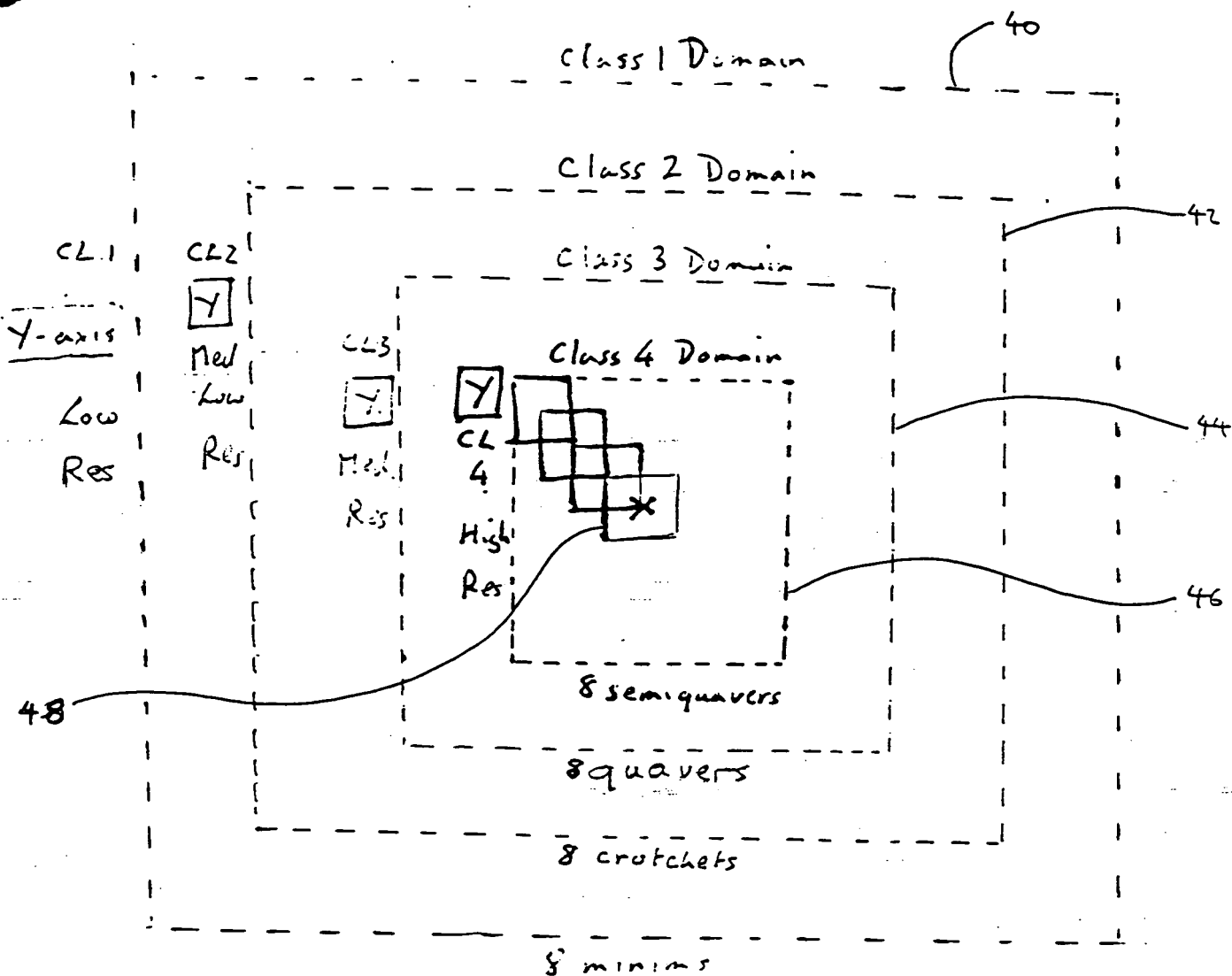


Fig. 4

This Page Blank (uspto)

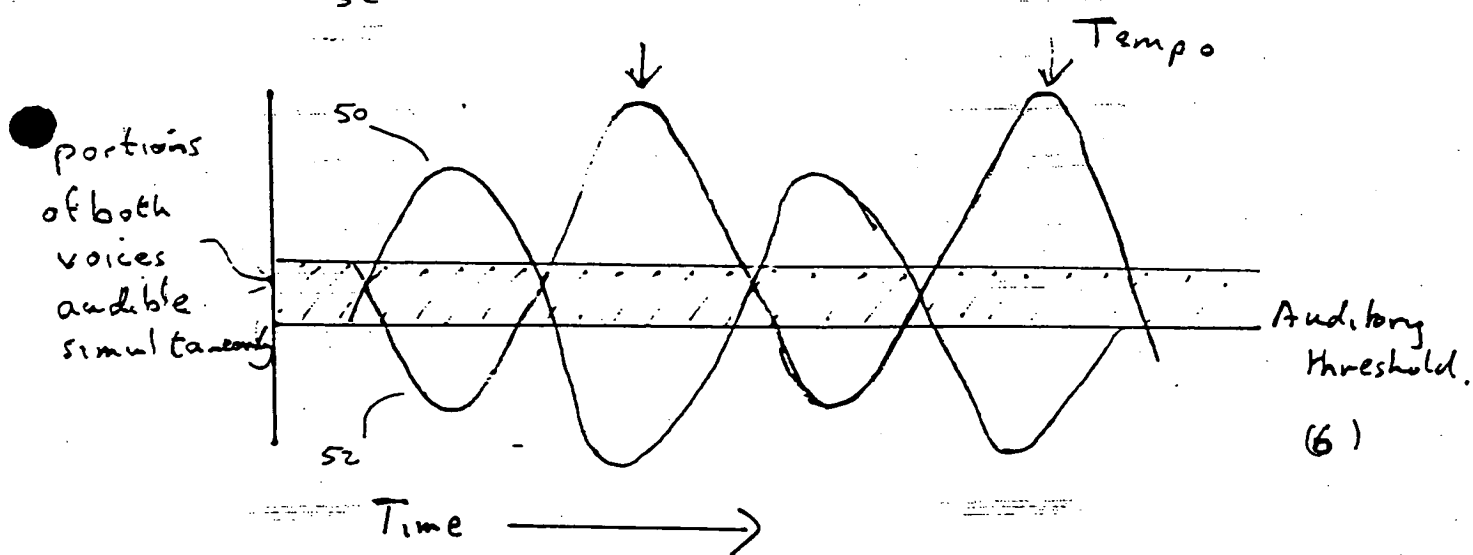
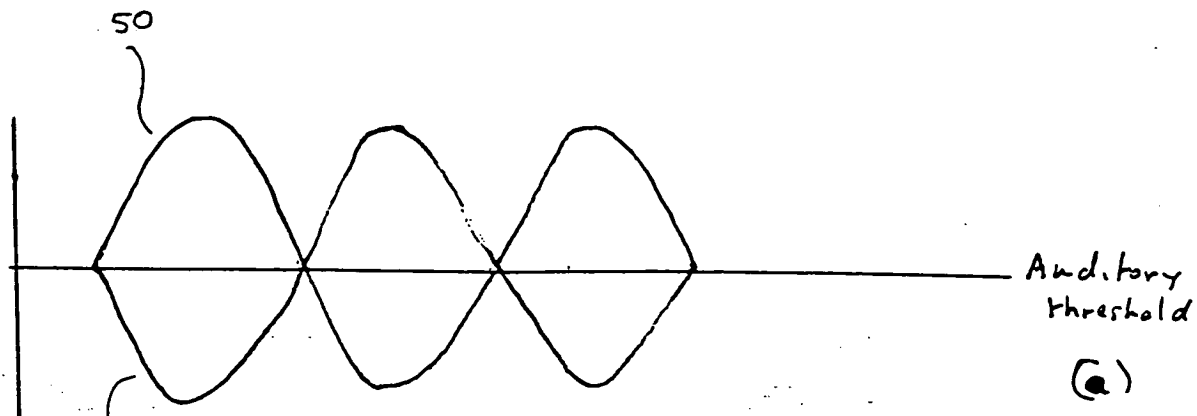


Fig. 5

PCT/GB 99/01500

13.5.99

McNeight + Lawrence

This Page Blank (uspto)